THE DEVELOPMENT AND COGNITIVE TESTING OF RACE AND ETHNIC ORIGIN QUESTIONS FOR THE YEAR 2000 DECENNIAL CENSUS[1]

Eleanor Gerber and Manuel de la Puente
U.S. Bureau of the Census [2]

ABSTRACT

In this paper we describe how cognitive interviewing techniques were used to develop race and ethnic origin questions for inclusion in two major Census Bureau tests conducted in preparation for the Year 2000 Census. These are the National Content Survey and the Race and Ethnic Targeted Test. These tests are part of the Census Bureau's comprehensive program of research and testing on race and ethnicity. Our cognitive research examined a number of issues to improve race and ethnic reporting. However, the scope of this paper is limited to two key issues with wide-ranging implications for the collection and reporting of race and ethnic data. These are: (a) providing respondents who identify with more than one race with the opportunity to report their full racial heritage on the census form and (b) providing Hispanic respondents with the option of reporting their Hispanic origin in a census question that combines both race and Hispanic origin.

Based on our cognitive research we conclude the following: First, developing race and ethnic origin questions for the decennial census entails a series of compromises and trade offs. These include, lack of space on the census form and diverse and often competing suggestions offered by advisory committees, stakeholders and the general public.

Second, there is no perfect or ideal way to ask survey respondents to report race and ethnic origin. A question which serves the needs of one segment of the population may not include the best way to word or format the question for another group in the population.

And third, respondents' prior experience with race and ethnic origin questions in surveys is an important determinant of how they interpret and respond to race and ethnic origin questions in self-administered survey and censuses. Most respondents approached our questions with a strong habit of response. For this reason, unexpected modifications such as, changing the instructions in the race question to allow respondents to check more than one racial origin, will go unnoticed by some respondents. However, if this, or other somewhat unexpected changes in race and ethnic reporting are routinely included on survey and census questionnaires, then these methods of reporting will become institutionalized or accepted over time. We believe that this

---

[1] Paper presented at the Bureau of the Census 1996 Annual Research Conference, March 17-19, 1996 in Arlington, Virginia. This paper is published in the proceedings of the Annual Research Conference.

[2] This paper reports the results of research undertaken by Census Bureau Staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

institutionalization will lead to greater awareness and use of the new reporting feature among the general population. Therefore, for example, if race questions eliciting more than one racial or ethnic background are routinely included on survey and census questionnaires, the population identifying with more than one race or origin will more than likely increase over time, demographic changes notwithstanding.

KEYWORDS

Race; Hispanic origin; Cognitive Research; Multiracial.

I INTRODUCTION

In 1996 the Census Bureau will conduct two major tests in preparation for the Year 2000 Census. These are the National Content Survey (NCS) scheduled for March 1996 and the Race and Ethnic Targeted Test (RAETT) planned for June 1996. These tests are part of the Census Bureau's comprehensive program of research and testing on race and ethnicity.[3] This program has two purposes. These are: (a) to determine the content and format of the race and ethnic questions to be included in the Year 2000 Census and (2) to provide information needed for the Office of Management and Budget's (OMB) review of its "Race and Ethnic Standards for Federal Statistics and Administrative Reporting" contained in Statistical Policy Directive No. 15 (OMB 1995).

Directive No. 15 was issued by OMB and adopted in 1977. This Directive provides the minimum classifications which all Federal agencies must use in recordkeeping, collecting, and presenting data on race and ethnicity in statistical activities and program administrative reporting. If OMB changes the Directive as a result of its review, then the changes would affect the race and ethnic questions included in the Year 2000 Census forms. Thus, the NCS and the RAETT are essential for evaluating how proposed changes in Directive No. 15, such as the inclusion of a multiracial category in the race question and using a combined race and Hispanic origin question, affect the quality and distribution of race and ethnic data.

The Census Bureau is not the only Federal agency examining the potential impact that changes in Directive No. 15 can have on the quality and utility of race and ethnic data. In order to facilitate and coordinate the collaboration and cooperation of Federal agencies with the review of Directive No. 15, the OMB established the Interagency Committee for the Review of the Racial and Ethnic Standards. This Committee was established in March 1994, and its members represent over 30 Federal agencies ranging from principal statistical agencies like the Census Bureau and the Bureau of Labor Statistics to monitoring and enforcement agencies, such as the Office of Civil Rights in the Department of Education and the Equal Employment Opportunity Commission.

---

[3] For details on the Census Bureau's program of research and testing for race and ethnic origin see McKenney et al. 1996.

This paper documents research using cognitive interviewing techniques to develop race and ethnic origin questions for inclusion in the two previously mentioned Census Bureau tests -- National Content Survey (NCS) and the Race and Ethnic Targeted Test (RAETT). The cognitive research reported in this paper was guided by (a) concerns about data quality that have emerged from Census Bureau evaluations of the race and Hispanic origin items in prior decennial censuses and (b) claims by stakeholders, federal and local government agencies and the general public that the current race and ethnic categories do not adequately represent the racial and ethnic diversity in this country.[4]

Our cognitive research examined a number of issues to improve race and ethnic reporting. These included new racial and ethnic terminology coming into popular use, new racial classifications, and the most appropriate sequencing of the race and Hispanic origin items on the census form. However, the scope of this paper is limited to two key issues with wide-ranging implications for the collection and reporting of race and ethnic data. These are: (a) providing respondents who identify with more than one race with the opportunity to report their full racial heritage on the census form and (b) providing Hispanic respondents with the option of reporting their Hispanic origin in a census question that combines both race and Hispanic origin.

This paper describes the context in which our research occurred by (a) summarizing key findings from Census Bureau evaluations of the race and Hispanic origin items; (b) discussing the demographic factors that have prompted some to call for a change in the categories currently used for collecting and reporting race and ethnic origin data; and (c) providing background for the two major issues examined in this paper. This is followed by a detailed discussion of our major findings. The last section offers our concluding remarks.

II CENSUS BUREAU EVALUATIONS

Race is a complex and difficult concept to measure. Bates et al. (1994) noted that census racial classifications (from 1790 to 1990) have changed over time reflecting demographic shifts as well as changes in how Americans view race. The authors also point out that racial classifications are a mixture of principles and criteria which include national origin, language, minority status and physical characteristics, among other factors. Crews and Bindon (1991) argue that race is a sociological construct with little correlation with detectable biological phenomena. Physical anthropologists have also stressed the inconsistencies of attempting to define race in biological terms (Gould 1981).

Cognitive research conducted to develop questions for the May 1995 Current Population Survey (CPS) supplement on race and ethnic origin revealed many survey respondents tend to use

[4] The race and ethnic origin questions tested in our cognitive research are also based on extensive consultation activities undertaken by the Census Bureau, including an international conference on race and ethnicity co-sponsored with Statistics Canada, congressional hearings, the National Academy of Sciences workshop on race and ethnicity, deliberations of the Federal Interagency Committee, the Research Subcommittee's research agenda, comments from public hearings, and additional consultation with Census Bureau advisory committees and a broad spectrum of racial and ethnic organizations.

the terms "race" and "ethnic origin" interchangeably, and they do not clearly distinguish between the two concepts. This research also showed that respondents find survey questions on race and ethnic origin sensitive, and for some, such as Hispanics and persons with more than one race in their background, difficult to answer (McKay and de la Puente 1995a; de la Puente and McKay 1995b).

Below is a summary of Census Bureau evaluations of the census race and Hispanic origin items in the context of the decennial census. These evaluations indicate that while overall the race and Hispanic origin items provide good quality data, there are areas where improvement in reporting is needed.

The Race and Hispanic Origin Questions on the Census Form

Two indicators of data quality used by the Census Bureau are consistency of response derived from Content Reinterview Surveys (CRS) and computer allocation rates. Inconsistencies occur when respondents provide different answers during the reinterview than on the census form. Computer allocation is required to impute values when respondents leave questions on the census form blank.

Although the 1980 and 1990 CRS showed that, in general, there is good response consistency for the race question; there are problems with specific racial categories. For example, the 1990 CRS revealed low response consistency for the "American Indian", "Other race" and "Other API" response categories when compared to the other race options contained in the race question (McKenney and Cresce 1992; McKenney et al. 1993).

Overall, the allocation rate for the 1990 race question was about 2 percent. This compares favorably to the other demographic questions on the census form, with the exception of Hispanic origin (see below). However, the 2.0 percent allocation rate for the race item in 1990 is higher than the 1.5 percent allocation rate for the same item in the 1980 census (Cresce et al 1992; U.S. General Accounting Office 1993).

The 1990 CRS revealed that the Hispanic item enjoys a relatively high consistency of response. The CRS showed that almost 90 percent of respondents who classified themselves of Hispanic origin in the census also did so in the reinterview. A closer look at these data show that most of the "switching" between Hispanic and non-Hispanic occurred in the "Other Hispanic" and the "Mexican" response categories. This pattern was also observed in the 1980 census (McKenney et al. 1993).

In 1990 the percent of consistent responses in the "Other Hispanic" category was 63 percent. This does not compare favorably with the 82 percent consistent response for each specific Hispanic origin group. However, the 63 percent rate for the "Other Hispanic" was an improvement from the 1980 rate of 55 percent. We believe that the addition of a write-in line for the "Other Hispanic" category accounts for the modest improvement from the 1980 rate. Nonetheless, the results still point to a response problem that needs to be addressed.

The allocation rate for the Hispanic origin question continues to present the Census Bureau with significant challenges. In 1990 the allocation rate for this item was 10 percent -- the highest of all the 100 percent items including race. In 1980 the allocation rate for the Hispanic item was much less -- 4.2 percent (Cresce et al 1992; U.S. General Accounting Office 1993). The relatively high allocation rate for the Hispanic origin item in 1990 is due, to a large extent, to the lower level of follow-up which occurred in 1990. However, the high level of nonresponse by non-Hispanics to the Hispanic origin item and the placement of the race item before the Hispanic origin item also contributed to the high allocation rate for the Hispanic origin item.

Despite the acceptable overall consistency for the race and Hispanic origin items and relatively low allocation rates associated with these items, there are reporting problems for some population groups. For example, problems with race reporting among Hispanics occur because current racial categories are incongruent with the way in which some Hispanics view race (e.g., see, de la Puente and McKay 1995b; Bates et al. 1996; Rodriguez 1992 and 1994). The differences are manifested in distinct categories and criteria for race.

In general, race in this country is viewed as an objective fixed characteristic of a person which is biologically inherited. Persons from Central and South America tend to define race as a continuum rather than as discrete categories, such as "White" and "Black." These cultures also tend to use more and different racial labels and criteria. Rodriguez (1992 and 1994) notes that in Latin America there is a relatively large number of racial terms for "intermediate" racial categories such as "mulatto" or "moreno." These cultural factors may create problems for recent immigrants to the United States in understanding the race and Hispanic origin items. Other Hispanics in this society offer the term "Hispanic" as a racial identity.

Given the incongruence just described, it is not surprising that the number of Hispanic origin persons classified as "other race" in the race question increased from 700,000 in 1970 to 5.8 million in 1980 and 9.7 million in 1990[5] (Cresce et al 1992). Census Bureau research also shows that in the 1990 census over 97 percent of the 10 million persons who reported in the "Other race" category are of Hispanic origin (U.S. Census Bureau 1992b). Not everyone who checks the "Other race" category fills in the write-in box. In 1990 there were about 2.5 million write-in responses to the "Other race" category. Census Bureau tabulations show that of these 2.5 million write-ins about 53 percent were from persons of Hispanic origin (Cresce et al. 1992).

In the 1990 census the race question preceded the Hispanic origin question on the census form. In the 1990 census, 373,100 persons who provided a Hispanic write-in response (such as "Mexican", "Puerto Rican" or "Spanish") in the race question did not respond to the Hispanic origin question. Cognitive research, as well as in-depth interviews and focus groups, with Hispanics of different national origins show that some Hispanics find the race and Hispanic origin questions redundant because these questions are viewed as asking for the same information (McKay and de la Puente 1995a; de la Puente and McKay 1995b; Kissam et al. 1993). These

[5] However, some of the increase in "other race" reporting by Hispanics may be due to the fact that in 1970 the Census Bureau began to conduct the census by self-enumeration rather than personal visit enumeration.

findings were confirmed in our research.

The cognitive research documented in this paper addresses some of these areas of concern. For instance, a combined race, Hispanic origin and ancestry question was designed in order to decrease the substantial number of Hispanics who use the "other race" category. (The question was also designed to improve the reporting of other ethnic and racial groups.)

The other areas of concern identified by Census Bureau evaluations such as the relatively high item non-response rate for the Hispanic origin item by non-Hispanics, the inconsistent reporting in the "American Indian", "Other race" and "Other API" categories as well as inconsistent reporting in the "Other Hispanic" category in the Hispanic origin item were also addressed in our cognitive research but are not reported in this paper.[6]

III THE CURRENT STANDARD FOR COLLECTION AND REPORTING OF RACE AND ETHNIC ORIGIN DATA AND THE CALL FOR CHANGE

The Current Standard

The Office of Management and Budget's (OMB) "Race and Ethnic Standards for Federal Statistics and Administrative Reporting" contained in Statistical Policy Directive No. 15 was adopted in 1977. Directive No. 15 specifies the collection and publication of data for four racial groups -- American Indian or Alaska Native, Asian or Pacific Islander, Black and White -- and one ethnic group, Hispanic origin (OMB 1995). To a large extent, the racial and ethnic categories specified in Directive No. 15 reflect the legislatively-based priorities for data on certain population groups and not efforts by population groups to be specifically identified.

For instance, collection of Hispanic origin data is required to implement the requirements of Public Law 94-311 of June 1976 which called for the collection, analysis and publication of social and economic data on persons of Spanish origin or descent. Data collected and reported in the categories provided in Directive No. 15 are also used for the enforcement of the Civil Rights Act of 1964, the Voting Rights Act of 1965 and the Equal Employment Opportunity Act of 1972 among other legislative requirements (OMB 1995; Evinger 1995).

Racial and Ethnic Diversity

The increase in racial and ethnic diversity in the United States has been well documented. Between the 1980 and 1990 decennial censuses, the rate of increase for Blacks was 13 percent; for American Indians, Eskimos and Aleuts; it was 38 percent; for Asians or Pacific Islanders; it was 108 percent. The rate of increase for Whites was only 6 percent (U.S. Census Bureau 1992b). According to recent population projections, in the next 50 years the White population is expected to decrease from 75 percent of the total population reported in the 1990 census to 59 percent. These projections imply that the relative sizes of minority populations will increase

[6] Our findings in these other areas will be documented in a separate Census Bureau report.

substantially. However, while the Hispanic and Asian populations are expected to increase, the Black population is expected to remain at about its current relative size. The following breakdown in 2040 are indicated by these projections: Hispanics 18 percent; Blacks 12 percent; Asian and Pacific Islanders 10 percent; and American Indian and Alaska Natives 1 percent (Edmonston and Passel 1994).

Census data show that the U.S. is more racially and ethnically diverse today than in the past. For example, members of ethnic/racial groups in 1900 totaled 10 million and increased to 20 million by 1960. Rapid growth occurred after 1960. The number of members of ethnic/racial groups went from 20 million in 1960 to 60 million in 1990. Immigration to the U.S. was a key element in this growth.

The Hispanic origin population also exhibited growth between the 1980 and 1990 censuses. Hispanics of Mexican origin, the largest of the Hispanic subgroups, totaled 13.5 million in 1990, an increase of 54 percent between 1980 and 1990. Similarly, Puerto Ricans, the second largest Hispanic subgroup, totaled 2.7 million in 1990, yielding a 35 percent increase since 1980. The Cuban origin population constituted 5 percent of all U.S. Hispanics in 1990 and showed a 30 percent increase between 1980 and 1990. Other Hispanics had a 67 percent increase between 1980 and 1990 (U.S. Census Bureau 1992b).

Another factor contributing to the increasing racial diversity in this country is the growth of interracial marriages and the increasing number of children in such families. For example, the number of interracial marriages involving at least one white partner in 1970 was approximately 150,000. The number of such marriages more than doubled to over 300,000 in 1980 and totaled 1.1 million by 1990. In 1970, the number of children living in families where one parent was White and the other non-White, (Black, American Indian, or Asian or Pacific Islander, and "other race") was 400,000. This number increased to 570,000 in 1980 and 1.5 million in 1990 (Bennett et al. 1995).

Call for Change and OMB's Response

In recent years Directive No. 15 has come under criticism from those who believe that the racial and ethnic categories in the Directive do not reflect the growing racial and ethnic diversity of the country. Some have called for the inclusion of a "multiracial" category for persons who identify with two or more races. Others have argued for the inclusion of new categories such as "Middle Easterner", "Arab American" and "Cape Verdean." Still others have proposed changing the names of existing racial and ethnic categories (such as "African American" to replace "Black" and the use of "Latino" instead of "Hispanic origin".) (OMB 1995.)

In response to this criticism OMB announced a review of the Directive in hearings held by the House Subcommittee on Census, Statistics, and Postal Personnel, (Evinger 1995) to address the measurement of race and ethnic origin in the decennial census.

The OMB's review process has included a workshop organized by the Committee on National Statistics (CNSTAT) of the National Academy of Sciences as well as public hearings and

solicitation of comments by federal agencies on their requirements for data on race and ethnic origin (OMB 1995).

To assist its review, OMB established the Interagency Committee for the Review of the Racial and Ethnic Standards in March 1994. Under the Committee's aegis, a Research Working Group, developed a research agenda for the review of the Directive (Research Working Group 1995). The research agenda outlined a number of research issues including the inclusion of a multiracial category in the race question and combining the race question and the Hispanic origin question into one question.

OMB's review process also included the publication of a Federal Register notice in June 1994 announcing the review of Directive No. 15 and asking for public comment on the adequacy of the race and ethnic standards in the Directive, the principles which should guide the review of Directive No. 15 (if a revision were to occur) and specific suggestions for change in the Directive (OMB 1994 and 1995).

Current statistical standards do not provide a separate category for multiracial persons to respond in. According to Directive No. 15 persons of mixed racial or ethnic origin should select one of the categories "...which most closely reflects the individual's recognition in his community..." (OMB 1995). A growing, but small proportion of the U.S. population who identify with more than one race resent having to report one race or ethnic origin. Some of these persons argue that the current Directive forces persons of mixed racial parentage to deny part of their racial heritage.

Two methods of collecting Hispanic origin data are noted by the current statistical guidelines. Directive No. 15 provides the option of collecting data on Hispanic origin either in a separate Hispanic origin question or in a question combining ethnic origin and race. If Hispanic origin data are collected in a combined question, the Directive states that the number of White and Black persons who are of Hispanic origin must be identifiable (OMB 1994).

Directive No. 15 notes that it is preferable to collect Hispanic origin data separately from race data in order to provide maximum flexibility (OMB 1994). The Census Bureau has collected race data since 1790. Beginning in 1970, a Hispanic origin question (separate and apart from the race question) appeared only on the 5-percent sample form. In the 1980 and 1990 censuses the Hispanic origin question was moved to the 100 percent form and was asked of everyone.

Federal agencies and many other stakeholders recommend a separate race question and a separate Hispanic origin question as the best way to obtain data on race groups and the Hispanic origin population. However, some organizations and researchers argue for a combined question as a means of reducing reporting problems.

IV METHODS

Phases of the Research

The cognitive research reported in this paper was designed to develop testing options for two major Census Bureau tests, the National Content Survey (NCS) and the Race and Ethnic Targeted Test (RAETT). The research was iterative, as new options were added for testing and as questions were revised on the basis of research findings. An attempt was made to recruit a wide variety of respondents from many different communities and in different geographic locations. The following briefly summarizes the phases of research and the mix of respondents in each phase:

- Research for the National Content Survey (NCS) focused on developing options for a multiracial category. The research occurred in two phases. A total of 32 respondents were interviewed, 14 in Phase I and 18 in Phase II. All cognitive interviews were conducted by Census Bureau social scientists from the Center for Survey Methods Research (CSMR). The 32 respondents who were interviewed for the NCS cognitive research included 7 White, 9 African American, 6 Hispanic, 3 Asian, and 7 Multiracial respondents.

- Research for Race and Ethnic Targeted Test (RAETT) added some additional options for multiracial reporting and tested options for including the term "Hispanic" in the answer categories for the race question. The research occurred in four phases. In Phase I, 37 cognitive interviews were conducted. These respondents included 4 White, 4 Black, 6 American Indian, 5 Asian, 8 Hispanic and 10 respondents who were themselves multiracial or who were the parents of multiracial children. Thirteen of these interviews were conducted by CSMR staff in the Washington, DC area and in Atlanta, GA. Outside researchers conducted the remaining 23 cognitive interviews under purchase orders in Riverside CA, Boston MA, Broward County FL, and Albuquerque NM.[7]

- Phases II, III, and IV of the RAETT cognitive research examined and refined format and wording options for the strategies tested in Phase I. In Phase II of cognitive research for the RAETT, a total of 12 cognitive interviews were conducted in Boston MA, Broward County FL, and Riverside CA, by non-Census Bureau social scientists. The respondents included 2 White, 2 West Indian, 2 Hispanic, 2 Asian or Pacific Islander, and 4 Multiracial respondents. In RAETT Phases III and IV, 5 and 6 cognitive interviews, respectively, were conducted by CSMR social scientists in the Washington, D.C. metropolitan area. The respondents in these two phases of the research included 4 White, 5 African American, 1 Asian and 1 Multiracial respondent.

- In all there were a total of 92 cognitive interviews conducted in all phases of the research.[8]

---

[7] We are indebted to Adalberto Aguirre, of the University of California, Riverside; Carol Lujan of Arizona State University; Peter Hainer of Curry College; and Judy Wingerd, research consultant in Ft. Lauderdale, Florida, for their contributions to this research.

[8] An additional 11 interviews were carried out by Census Bureau staff in Alaska. These interviews concentrated on terminological issues rather than the main research goals described in this paper. Their results have been reported separately. A focus group and a series of classroom experiments were also undertaken. Findings from these efforts will be reported elsewhere.

Cognitive Interviews

Cognitive interviewing for both the NCS and RAETT self-administered forms employed the following procedures:

- *Experimental forms.* Prior to interviewing, the experimental questions were incorporated into forms that were adapted from the 1995 Census Test form. This was done to ensure that the questions were understood as much as possible within a census context.

- *Interview protocols.* Detailed protocols were written to ensure comparability in the interviews of various researchers. Protocols covered the kinds of probes to use and the observational material necessary to understand respondents' progress through the form.

A conference call was conducted with non-Census Bureau social scientists participating in the research to ensure that the interview protocols were administered in the same way in all cognitive interviews.

Respondents were probed on all questions, not just race/ethnicity. This helped to acquaint respondents, and also served to de-emphasize our interest in a potentially sensitive set of questions. Both concurrent think-aloud probes and retrospective probes were used (Forsyth and Lessler, 1991). Concurrent probes such as "What are you thinking" were used throughout the interview, to encourage respondents to verbalize their reactions without delay. However, it was important not to interrupt respondents in certain critical sequences of questions, and interviewers were instructed not to probe until respondents had completed their answers to these sequences. In particular, respondents were allowed to fill out the entire sequence of the ethnicity and race questions prior to probing on the meaning of specific terms and phrases. The probes encouraged respondents to give reasons for their particular answers. They were also asked to explain the meaning they attached to the questions, by providing paraphrases of them. ("Can you tell me in your own words what this question is asking?") Specific words or phrases were also probed. All respondents were asked about the meaning of the term "race" and the term "Hispanic". Other terms relevant to particular questions were probed, such as "origin", "ancestry", "national origin", "multiracial", "biracial", "ethnicity" and "ethnic group." They were also asked to explain the meaning of the specific categories that they chose in their answers. The interviewers were also instructed to create other probes, as needed, to better understand terms and phrases which were offered spontaneously by respondents.

- *Conduct of the interview.* Respondents were asked to read aloud as they filled out the census form. This reading was not interrupted by interviewer questions. Interviewers were asked to specifically note any abbreviations or changes in the reading of the questions. This helped to identify parts of questions which were difficult to read or to which respondents did not attend. Debriefing questions were asked at the end of the interview about the difficulty and sensitivity of the questions. In the debriefing section, we

also elicited the race of the respondents' parents.

Recruiting

It was central to our aims to understand whether the questions we designed worked, for people of varied backgrounds. As a result, we recruited respondents from a number of races and ethnic backgrounds including White, Black, Hispanic, American Indian, Alaska Native, Asian and Pacific Islanders of different national origins, Hispanics of different national origins, and persons with parents of different races or ethnic origins (multiracial respondents). Our non-Census Bureau social scientists recruited their respondents in accordance with guidelines provided by the Census Bureau.

We employed three recruiting strategies. First, we recruited respondents (especially multiracials) through contact persons who knew respondent's racial and ethnic background. Second, respondents were chosen from CSMR's database of local area respondents who had participated in other unrelated research.

Respondents were told that they would be helping the Census Bureau test experimental census forms in preparation for the year 2000 census. They were not told that we were specifically interested in how they report their racial and ethnic background and that of other members in their households. We did not wish to alert respondents prior to interviewing about special features of the questions.

The recruiting of multiracial respondents proved particularly challenging. Although we know ahead of time, either from the CSMR database or personal contact, that people had more than one race in their background. The majority of these persons did not choose to self-identify as multiracial during the cognitive interview. Although such respondents often freely discussed the various races in their backgrounds with the interviewer, their choice was to report only one race on the census form. We encountered this pattern with potentially multiracial respondents (or proxies for them) who reported themselves (or family members) as Black, Asian, Indian, Pacific Islander, and Eskimo. In some instances these persons had parents of different races. In other cases, the multiracial aspect of their background was further back in their heritage.

Although such respondents provided insight into the cultural, sociological and psychological definition of race, they did not meet the requirement for respondents who would opt to identify themselves (or other household members) as multiracial on the experimental forms.

In order to find respondents who would be more likely to use multiracial response options, additional respondents were recruited through an organization with a strong political interest in creating a multiracial category. They included both multiracial persons and parents of multiracial children. The seven respondents with a self-selected interest in this topic reacted to our experimental forms quite differently than most respondents recruited using our ordinary recruitment methods. They had formed opinions about the multiracial category and were primed to look for it among the categories. Respondents who were not recruited in this way were more likely to be unaware of the multiracial category when it was offered to them. Although these two

"types" of respondent intergrade, this indicates that the recruiting of multiracial respondents is a critical element of research on this issue. We found that a mix of highly-interested and less-interested respondents will probably provide the most balanced results.

V FINDINGS: MULTIRACIAL OPTIONS

We developed two different response options for race reporting by respondents who identify with more than one race. The first option offers a separate category for reporting multiracial background. The second response option allows respondents to select more than one response category in the race question. The following outlines the development of each of these options.

Option One: "Multiracial" as a Response Category.

Background

This option is supported by a number of advocacy groups of multiracial persons and the parents of multiracial children, who view this as an emerging form of racial identification in our society. In Congressional hearings and letters to the OMB and the Census Bureau, these individuals have specifically requested the category "Multiracial" to appear as one of the response options in the race question.

Research on developing a "multiracial" category extended through both rounds of testing for the NCS and the first phase of cognitive research for the RAETT. In all, 37 cognitive interviews were carried out with 7 White, 5 African American, 4 American Indian, 7 Asian and Pacific Islander, 6 Hispanic and 8 Multiracial respondents.

Placement of the "Multiracial" Category

In the first round of cognitive testing for the NCS, determining where to best place the category "Multirace" was a challenge. (See Attachments A and B) Originally, this category was placed between "Other Asian and Pacific Islander" and "Some other race" categories. As shown in Attachment A, both of those categories have write-in lines, as did the category "multirace". This was problematic from a forms design perspective because it was difficult for respondents to find and distinguish the "Multirace" category from the other categories that also had write-in lines.

In NCS Phase I, most respondents did not notice the "Multirace" category; even those for whom it might have been relevant did not notice this feature of the question. Other respondents confused the closely placed write-in boxes and used them for unintended purposes, such as writing an entry intended for "some other race" into the "Multiracial" write-in box.

In NCS Phase II, the write-in box for the multiracial category was placed last, that is, after the write-in box for "some other race." (See Attachment B) It has retained this position

throughout the remaining cognitive testing and in the RAETT.

The Number of Write-in Boxes

Write-in boxes were provided to enable multiracial respondents to specify the races with which they identify. When only one box was provided (as in Attachment A), some respondents thought that they were only asked to write one racial identification. This effect may have been exaggerated by segmentation of the write-in boxes (not visible in Attachment A), in preparation for data capture using electronic imaging. (Segmentation splits the horizontal answer box with pale vertical lines, so that respondents will print letters in discrete spaces.) Respondents could see that a limited number of letters were expected, with no breaks indicating more than one word. These visual cues were interpreted to mean that only one race was being elicited. This conveyed a contradictory message to most respondents since they had checked a box indicating that they were multiracial, and yet they were being asked to enter only one race in the write-in box.

Some multiracial respondents were opposed to having only one write-in line for the multiracial category. These respondents erroneously believed the single race write-in entry would be used to assign them a single race rather than categorize their response as "Multiracial." This belief was characteristic of respondents, who view "Multiracial" as a new and emerging social identity, and have a high degree of interest in the topic. Such respondents are particularly sensitive to the suggestion that they are being required to ignore or deny part of their backgrounds.

Two write-in lines for the multiracial category are desirable for respondents, but there are potential problems with this strategy. Too many write-in lines in the entire question can confuse respondents by visually breaking up the question. In our interviews we noticed that some respondents regarded the write-in line for "enrolled or principal tribe" directly under the American Indian response category as the end of the race question. Initially, these respondents had difficulty finding additional race categories and occasionally commented that the response categories that come after the "American Indian" category constitute a separate question. (This write-in line directly after the American Indian category is included in all forms because collecting tribal data is mandated by legislation.) However, the visual appearance of multiple write-in lines and limitations of space make it difficult to add additional write-in lines to the question. It was also thought that there might be negative reaction to the use of so much space for a multiracial response option, since it is applicable to only a relatively small segment of the general population.[9]

We tested an experimental form in which the multiracial category shared a write-in box with the "some other race" category. (See attachment K) However, this design proved difficult for respondents to interpret and was consequently dropped. Furthermore, we encountered some reporting problems in the 1990 census when "other API" and "other race" shared a single write-in box.

---

[9] According to the May 1995 Current Population Survey supplement on race and ethnic origin, approximately 1.6 percent of respondents self-identified as multiracials when this category was offered on the race question (Tucker et al 1996).

Terminology

The initial race question in the NCS Phase I testing (see Attachment A) used the term "multirace." This was changed to "Multiracial or Biracial" (see Attachment B) because some respondents indicated that "multirace" meant "more than two races", while "biracial" meant "two races." (Most other respondents thought that multiracial could apply to two races as well as more than two races.)

Respondent reactions to the "multiracial or biracial" category

The term "multiracial or biracial" is generally understandable to respondents regardless of racial or ethnic background. However, some multiracial individuals are hesitant to supply a write-in race, particularly in formats where there is only one write-in line, because they believe that they will be "coded back" to a single race. Some even prefer versions with no write-in lines.

The presence of the category "Multiracial or Biracial" did not appear to change the reporting of respondents who knew of more than one race in their backgrounds but self-identified as only one race. As we have previously noted, most persons who were recruited as multiracial (because they mentioned several races in their family history) chose not to report in this category during the cognitive interview. Although these respondents did not hesitate to discuss the complexities of their racial and ethnic heritage with the interviewer, encountering the "multiracial or biracial" category did not cue them to think about changing their response on the census form.

Option 2: Choosing More Than One Race

Background

Census forms containing the option of selecting more than one race (rather than a multiracial response category) have the advantage of preserving detailed data about racial affiliation which could be difficult to capture for the option discussed above. However, race data reported in a question which allows for more than one response can present difficulties during tabulation and analysis. These issues, however, were not within the scope of our research.

In developing a race question that provides respondents with the option of selecting more than one race, we were faced with formulating neutral or unbiased instruction for the respondent. On the one hand, it was important not to give the impression that respondents were expected or required to mark more than one race. Yet on the other hand, we wanted to convey to respondents that the question provided the option of reporting more than one race.

Research to develop the option of choosing more than one race included 44 cognitive interviews. Eight white, 12 African American, 2 American Indian, 5 Asian and Pacific Islander, 6 Hispanic and 11 Multiracial respondents were interviewed.

Version 1: Follow-up box

Early in our cognitive testing, we developed a question that enabled respondents to specify all applicable races but then requested that respondents specify the race with which he or she most identifies. We viewed this as a way of dealing with potential problems in statistical analysis arising from multiple responses for each respondent. Without a single race reported, allocation rules and statistical analysis of multiple responses would be difficult and very challenging.

In the first version which was cognitively tested, a complex instruction following the race question informed multiracial respondents about the new response option and the follow-up box (See Attachment C). This question and instruction read:

> ***What is this person's race?*** *Mark X ONE box for the race that the person considers himself/herself to be. Persons who identify with more than one race may mark more than one box and* ***write the race they most identify with in Box A below.***

Our cognitive interviews revealed that respondents could not absorb nor understand the complex instruction which accompanied this version of the race question. The instructions required respondents to process "race" three separate times. (Check the ONE race you consider yourself to be; if you identify with more than one race, you can check more than one; and write the race you most identify with in Box A.) Two of these instructions (checking one and checking more than one) could be viewed as mutually contradictory. The question was conceptually confusing to respondents.

The format was also problematic in other ways. First, respondents did not spontaneously find the Box A, even when they had completely read the instruction. Second, since Box A was directly after the "Some other race" write-in box and did not have any instruction directly above it, some respondents used it as the write-in box for "Some other race."

Because of these conceptual and formatting problems, the "Box A" version of the question was substantially revised for NCS Phase II. To facilitate respondent understanding, the question was reformatted as a two part question (See Attachment D) The instruction to mark the race (or races) was placed in Part A of the question. Part B asked the race with which the multiracial respondent most identified. Both parts are presented below:

> *5a.* ***What is this person's race?*** *Mark the box (or boxes) for the race (or races) that this person considers himself/herself to be.*
>
> *5b. If more than one box is marked in 5a, what race does this person MOST identify with?*

The instruction in 5a. was thought to be simpler than the former version, since it presented the entire instruction in a single sentence and did not give contradictory instructions.

In this format all but one respondent was able to spontaneously find the follow-up box.

They understood what it meant. However, they frequently were not aware of the feature until they encountered the follow-up box. In other words, the plurals contained in the parentheses in the instruction in 5a. did not provide enough information for respondents to comprehend the new question feature. Even when they read them fully, they gave no evidence of noticing the feature until they encountered the instruction in Part B of the question. At this point most respondents glanced up and remarked something like, "Oh, I guess you could give more than one race." Occasionally, they apologized for not "reading carefully" earlier, although in fact they had read the complete instruction in the first part of the question.

Space considerations: Dropping the Follow-up box

The cognitive interviews for NCS Phases I and II were conducted using booklet-style census questionnaires. These booklets contain half-page "person boxes" for the recording of demographic information about household members listed on the roster. Space in these boxes is limited. With a separate Hispanic origin question and a follow-up box in the race question, the race and ethnicity sequence takes up more space than it has been allotted in the past. Although the follow-up question itself does not take much space, the column allotted to the race and ethnicity questions is quite full. In fact, our tested version of the follow-up box only fit into the column because we were able to abbreviate the relationship question and move the Hispanic origin question to another column. Since cognitive testing showed that some respondents who identified with more than one race objected to specifying a single race in the follow-up box or did not provide a meaningful response, and because space was limited, this format was dropped.

Instructions for Choosing More Than One Race

As we have seen, the follow-up box provided a necessary cue for respondents, who often did not notice the "choose more than one race" feature until they encountered the second part of the race question described above. Without the follow-up box, respondents have to rely entirely on the instruction provided in the race question for this information.

The Census Bureau had never before used a race question in any survey or census with the "choose more than one race" feature. In addition to the instructions discussed above, we developed and tested a variety of respondent instructions to communicate this feature. Our objectives were to clarify and simplify the instructions and to formulate a neutral or unbiased instruction for the respondent.

The following versions of the instruction were tried during our cognitive testing:

(1) *Mark the box for the race that the person considers himself/herself to be. You may mark more than one box.* (See Attachment E)

(2) *Mark one or more races to indicate what this person considers himself/herself to be.* (See Attachment F)

(3) ***Mark one or more boxes*** *to indicate what this person considers himself/herself to*

be. (See Attachment G)

(4) ***Mark all that apply.*** *Mark the race(s) that this person considers himself/herself to be.* (See Attachment H)

Although we were unable to find a "perfect" instruction, we were able to develop a workable instruction, as we discuss later in this section. Respondents frequently skipped the instruction altogether. Many others read the instruction, but did not absorb it; that is, they gave no spontaneous indication of understanding that they had the option of marking more than one race. When the option was pointed out by the cognitive interviewer, respondents claimed that they had no memory of having read the words and assumed that they had skipped over them. This behavior was characteristic of respondents who were multiracial (or who had children or other household members who were multiracial) as well as respondents for whom the issue was less salient.

We believe that two factors account for the fact that many respondents did not notice the new instruction. These factors arise more from habits acquired by respondents' experience with filling out forms than from the design features of the census form.

First, survey respondents are accustomed to responding to race questions by selecting one racial group.[10] For example, a number of respondents commented that they "always" mark a particular category or respond in a certain way. They indicated that they know what they will mark as soon as they see that the question is asking for race. As one respondent put it, "I've done lots of these before. I do it as quickly as possible." We found that some respondents hurry through the race question because the question makes them feel uncomfortable or uneasy. Such respondents are not looking for new information, but rather they are looking for an easy way to complete the task with as little involvement as possible.[11]

The second factor is that survey respondents tend to look more closely at the response categories offered in the race question than they do at the race question itself or the accompanying instructions. The response categories seem to provide respondents with much of the needed information used to formulate a response. This is indicated by the fact that many respondents read a few of the answer categories before they were certain of what was meant by "race." This suggests that respondents have pre-set opinions about the response categories, drawing their attention away from the instruction. For example, one respondent was eager to tell us, "I see you have African American now. That's good." In this typical case, the respondent skipped from the word "race" directly to the answer categories, without lingering on the instruction.

---

[10] There are exceptions. Some school districts in selected states have added the category "Multiracial" to the race question but for the most part survey respondents are asked to select one racial group from a list of options or to write in one race.

[11] This finding is not unique to this research. In a separate research project undertaken to develop race and ethnic origin questions for the May 1995 CPS supplement on race and ethnic origins, researchers also found that some respondents experience discomfort when asked to report their race or ethnic origin or the race and ethnic origin of other members of their households (McKay and de la Puente 1995a; de la Puente and McKay 1995b).

In the context of a combined race/Hispanic origin/ancestry question, the instruction (using version 4, above) was somewhat easier for respondents to notice. However, it is important to keep in mind that the race/Hispanic origin/ancestry question has non-traditional format and wording; and, therefore, was less familiar to respondents. (See Part IV below for details on how the race/Hispanic origin/ancestry question was developed.) In this unexpected format, about half of the respondents noticed the new option for reporting more than one race.

Respondent reactions to the option of choosing more than one race

Respondents who identify with more than one race have varied reactions to the "choose more than one race" option. Some prefer to record all races and ethnic origins in their backgrounds. Generally, these respondents had favorable opinions of this option because it does not require them to "deny" any part of their racial and ethnic heritage. In fact, some respondents told us that they often check more than one racial category even though the question does not have this feature. Another pattern which is evident among multiracial respondents with no special interest in the topic is to mark off only one race. This may be done because it is "easier" or because the respondent believes that he or she is perceived by society as being a member of only one category.

There are, however, respondents who are committed to the "multiracial" category and see it as an emerging new identity. As noted earlier, most of these respondents were recruited by an advocacy group who is lobbying for the inclusion of a "multiracial" category on the census questionnaire and on all other forms which ask for racial or ethnic identity. These respondents do not favor the "choose more than one race" approach because the response option reports diverse aspects of a multiracial heritage rather than a unitary "multiracial" identity. Some of these respondents are also concerned that multiple reporting may be bureaucratically coded as a single race, despite the respondent's intention.

Evaluation of Options I and II

We believe that our cognitive research identified workable versions of option I and option II from the standpoint of both respondent understanding and respondent acceptability. Because of the small number of interviews and the means of selecting respondents, we are not able to judge which option will perform best in the population at large. The NCS and RAETT will provide the large scale testing required to determine data quality using such indicators as consistency of response and item non-response.

In terms of acceptability to respondents, some respondents who identify with more than one race or ethnic origin prefer to check off their multiple origins on the census form, while others clearly prefer to check a "multiracial" category. However, our findings also show that some respondents with multiple races or ethnic origins in their background prefer to specify only one race or ethnic origin even if they are aware of the options to report more than one race or origin.

VI FINDINGS: THE COMBINED RACE/HISPANIC ORIGIN/ANCESTRY QUESTION

Background

As noted earlier and as demonstrated by other research, some respondents, particularly Hispanics, view Hispanic origin as a race rather than an ethnic group (for example see, Rodriguez 1992; Kissam et al. 1993; de la Puente and McKay 1995a). Earlier we discussed Census Bureau studies showing the impact separate questions for race and Hispanic origin can have on data quality (for example see, Bates et al. 1994). To address some of these concerns and to be responsive to those who have called for a combined race and Hispanic origin question, we developed a single question which would include "Hispanic" among the race categories.

The design of our combined question reflects constraints dictated by the limited amount of space available on the census form as well as form design issues which can affect response. For instance, the combined question does not list either the specific Asian and Pacific Islander subgroups or the specific Hispanic origin subgroups, partly because of limited space on the census form. In order to compensate for this feature and obtain information on other groups such as Arab American, French Creole and Cape Verdean, we also incorporated ancestry into the combined question.

For the RAETT, we developed and tested two versions of a combined race/Hispanic origin/ancestry question. Both versions also provide reporting options for respondents who have more than one race or ethnic origin in their background. The first version includes the category "Multiracial or Biracial" and the second version provides the respondent with the option of selecting more than one race or ethnic origin.

Research for this reporting option included 12 cognitive interviews, including 2 White, 3 African American, 2 Asian and Pacific Islander, 2 Hispanic and 3 Multiracial respondents.

Conceptual Issues: Question Wording

Since the combined question is intended to elicit not only race but Hispanic origin and 0ancestry, we revised the wording of the question. As a result of cognitive testing, the question and the associated instructions have undergone a number of changes, which we describe below.

As originally authored, four relevant concepts were included in the question, along with a complex instruction. The question read:

> ***"What is this person's race, ancestry, ethnic group or national origin?*** *Mark X one box for the race or ethnic group with which this person most closely identifies himself/herself. Print a more specific ancestry, ethnic group or national origin in the space provided below."*

Each response category was followed by an instruction to "print ancestry below" and a list

of examples relevant to that particular race. The follow-up box was preceded by the following instruction:

> *"Print an ancestry, ethnic group or national origin. (For example, British, Cambodian, Cape Verdean, Dominican, Ecuadorian, Haitian, Cajun, French-Canadian, Irish, Jamaican, Lebanese, Nigerian, Polish, Russian, Slovak, Taiwanese, Ukrainian, etc.) If American Indian, print name of principal or enrolled tribe."* (See Attachment I)

We found a number of significant problems with this question. First, the question and instructions contained far too much text. As a result the question resembled a paragraph rather than a sequence of two questions. Second, the complete "text" required more space than was available on the census form. In addition, the question presented four potentially different concepts in a variety of combinations.

Prior to testing, we attempted to address these concerns by splitting the combined question into two parts in order to make it easier for respondents to process the information presented. In the revised version, all ancestry examples were removed from the race part of the question. The question and instructions were also shortened. (See Attachment J)

The Concepts of "Race", "Ethnic group", "Ancestry", "National origin" and "Ethnic origin"

Much of our research pertaining to the combined question was spent on finding the appropriate set of concepts to use in the question wording. Although the inclusion of "Hispanic" in the race question was in part motivated by evidence that Hispanic is regarded as a race, the question wording preserves the concept of ethnicity (as stipulated in Directive 15.) Therefore, in the first question tested, the wording of the question stem (See Attachment J) read:

What is this person's race or ethnic group?

The initial wording of Part B of the combined question was:

What is this person's ancestry or national origin?

Successive stages of cognitive testing provided us with greater understanding of how respondents interpreted "race", "ancestry", "ethnic origin" and "national origin" in the context of the combined question. Respondents' interpretations of these terms led us to modify question wording and terminology, as described below.

Race

We found that the term "race" is generally understandable to respondents in the context of the questions we tested. "Race" was sometimes defined in terms of skin color. Although for some respondents, there is hesitation to do so, perhaps because it may not be viewed as

"politically correct."[12] Typically, respondents defined "race" in terms of group membership or background. Terms such as "what you are," "your people" and "where your family comes from" were frequently used by respondents to define race. Often, respondents used the list of racial categories in the question to define race. Statements such as: "What you are, you know, white, black, Hispanic, whatever" were common.[13] The term "race" was the most important conceptual cue for respondents in the first part of the question, since they generally paid little attention to the term "ethnic group."

The Hispanic response category used in the question was "Hispanic, Latino or Spanish origin." We found that respondents had no difficulty in understanding these terms, although some non-Hispanics were unfamiliar with the term "Latino." Respondents usually defined the category in terms of Central and South American countries of origin.

Ethnic Group

Most respondents recognized the term "ethnic group" but did not distinguish it from the concept of "race." (Those who saw it as a distinct concept tended to be college educated.) In the context of the combined question, "ethnic group" was viewed as almost completely identical to the concept of "race." When probed about the meaning of "ethnic group," respondents would indicate that it was "the same thing" as race. In fact, both terms appear to be aspects of the same global concept. Respondents often coined the term "ethnic race" during our discussions. We substituted the term "origin" for "ethnic group" in order to acknowledge that some respondents may view Hispanic origin as distinct from race. Because the Hispanic origin response category in the combined question is worded as "Hispanic, Latino or Spanish origin," it made sense to use the term "origin" along with "race" in the stem of the question. Moreover, "origin" also reflected respondents' understanding of "race" in terms of family and group origins. The question wording tested in RAETT Phases I and II and used in the RAETT is:

What is this person's race or origin? (See Attachment K)

Ancestry

We found that most respondents recognized and understood the term "ancestry." There were a few respondents whose native language was Spanish who did not recognize this term. However, the examples provided served as a cue which made evident the intent of the question.

Respondents often defined "ancestry" by using the words "ancestors" or "family background" in their definition. "Ancestry" also has geographical connotations for most respondents as evidenced by definitions which included some version of the statement, "That's where your people are from." In their definitions, some respondents tended to refer to the

---

[12] Some respondents expressed the awareness that skin color may not match an individual's self-perception of race.

[13] These findings are in line with findings from the cognitive work conducted for the May 1995 CPS supplement on race and ethnic origin even though CPS supplement cognitive research was conducted in order to develop a CATI instrument and not a paper and pencil instrument like the census form (see McKay and de la Puente 1995a; de la Puente and McKay 1995b).

country of origin of their relatives. We also found that in debriefing respondents many freely offered a complete run down of all they knew about their origins including percentages (such as "50 percent Irish and 50 percent English").

National Origin

"National origin" was originally included in the wording of the question because prior research indicated that Asian and Hispanic immigrants primarily identify with their countries of origin (McKay and de la Puente 1995a; de la Puente and McKay 1995b). Although our current research bore out this finding, the introduction of the term "national origin" into the ancestry follow-up question (See Attachments J and K) created difficulties for many respondents. This occurred because "national origin" is defined in terms of country of birth by most people. (It is occasionally defined in terms of citizenship rather than place of birth.) As a result, for some respondents, "ancestry" and "national origin" are two quite different concepts. For example, White respondents often see their "ancestry" as a European country but their "national origin" as "America." This is also the case for Hispanic and Asian immigrants whose children are born in this country or who identify themselves as "American" because they are citizens. Complex migration patterns could also suggest contradictory answers to the question, as they did to an Asian Indian born in Guyana.

Therefore the inclusion of both concepts in the ancestry follow-up question was problematic. Since our interest was ancestry rather than country of origin, we dropped the term "national origin." The term "ethnic group" replaced "national origin" in Part B of the combined question. (See Attachment L) Although we did not cognitively test the combination of "ancestry and ethnic group", prior research has shown that most respondents understand the term when it is used separately (McKay and de la Puente 1995a; de la Puente and McKay 1995b).

Part B of the combined question now reads:

What is this person's ancestry or ethnic group? (See Attachment L.)

Example Effect

Another conceptual issue which arose during cognitive testing were example effects in the ancestry question. This phenomenon has been well documented by others (e.g. see Farley 1993).

In early versions of the combined question, we found evidence that some respondents interpreted the examples as a list from which they must make their selection. (See Attachment J.) We therefore made the list of examples shorter and added the phrase "or any other ancestry." (See Attachment L.) This modification seemed to be effective in relating to respondents that the list of examples was not a closed list of options.

The number of examples in the combined question used in the RAETT is now nearly the

same as it was in the first version of the question which was tested. We added to the list because we think certain groups might need specific guidance in the question. The content of the list of examples was also changed during the testing. This was done on the advice of an expert panel convened by the Census Bureau specifically to obtain advice on the NCS and the RAETT.[14]

Format: Multiple ancestry write-in boxes vs. a single follow-up box

The idea of using separate ancestry write-in boxes associated with each race group was recommended for testing by our expert panel. We, therefore, cognitively tested multiple write-ins. (See Attachment M.) We found that this format of the combined question was not conceptually clear to respondents. The only reference to "ancestry" in this format was contained in an instruction placed after the category name. (For example, "White - Print ancestry.") It was, therefore, harder for respondents to find and interpret the "ancestry" instruction, although most of them did provide ancestry write-ins. Another problem with this form of the combined format was that some respondents provided a write-in response but did not check the boxes. This was seen as a potential problem for data capture.

Based on these findings, we developed a combined question with two parts. Part A asks for "race" and "origin" and Part B elicits "ancestry" and "ethnic origin." This feature was retained for the combined questions which appear on the RAETT.

Number of write-in lines for ancestry follow-up question

On the 1990 decennial sample questionnaire, the ancestry question provided only a single write-in space. Despite this, many respondents used the available space to write in more than one ancestry. Segmentation of the answer boxes, which will be necessary for imaging purposes, clearly limits the number of spaces available. During cognitive interviews some respondents commented that the segmented spaces did not provide them with sufficient room to fully express their ethnic heritage. Further, since we are depending on Part B of the combined question to capture Hispanic subgroups, Asian subgroups, other race write-ins and multiracial write-ins we redesigned the combined question and added two write-in lines to Part B. (See Attachment L)[15]

Respondent reactions to the combined race/Hispanic origin/ancestry question

Respondents' reaction to the combined question was generally positive. They viewed the question as "fair" since it treats all groups in the same way. In testing of options which separate the race and ethnicity questions, we found that, for some respondents, a separate Hispanic origin

---

[14] Expert panel members include: Mary C. Waters (Harvard University), Matthew Snipp (University of Wisconsin), Reynolds Farley (University of Michigan), Juanita Tomayo Lott (Lott and Associates), John A. Garcia (The University of Arizona), and Lawrence Hirschfeld (University of Michigan).

[15] One version of the combined question provided respondents with the option of checking more than one race or origin from the list provided. Earlier we discussed how the instruction to this type of question was developed. (See Attachment G.)

question is not popular. Some have asked why Hispanics are signaled out. Others expressed the opinion that Hispanics are getting preferential treatment.[16] We did not encounter such reactions with the combined question. (However, it should be noted that these negative reactions occurred in the context of a cognitive interview and no respondent refused to fill out the census form because Hispanics had a separate question.)

In addition, some respondents viewed the combined question in a positive light because the question provided the opportunity for them to express their full racial and ethnic heritage. (They probably feel this way because of their sense that ancestry information is somehow more detailed and "personal" than information about race.) From our cognitive interviews we sensed that respondents tend to be less negative about the combined question than about versions which separate the race and ethnicity questions.

We found that the combined question worked well in eliciting subgroup information for Hispanics and Asian and Pacific Islanders.[17] That is to say, respondents easily provided ancestry write-ins which indicated countries of origin, such as "Peru" or "Vietnam." The intent of the question was clear even in cases where the family had a complex history of migration. For example, one Asian Indian respondent who had been born in Guyana understood the question as asking for his Asian Indian ancestry. We found very few respondents who showed evidence of a need for separate nationality listings. When they indicated a desire to include their own groups' names, it tended to be in the context of seeing other closely related nationalities on the form. For example, one Laotian respondent told us he didn't think it was necessary for "Laotian" to be on the form, but it ought to be there if we included "Vietnamese."

Despite its general acceptability, the combined question presents conceptual difficulties for some respondents, although no misreporting was observed in our cognitive testing. For example, American Indians are asked to provide their tribal affiliation after they check the American Indian category on the form. When these respondents were then asked to provide information about their ancestry or ethnic origin in Part B of the combined question, the information requested appeared redundant. However, despite this observation American Indian respondents did provide their ancestry information in Part B of the combined question.

African American also found the combined question somewhat redundant. To most respondents, ancestry refers to countries from which their ancestors migrated to the United States. African Americans were often not able to provide this information. In choosing a write-in for the ancestry box, most African American respondents relied on the inclusion of "African American" in the list of examples. It took some of these respondents a few minutes to find this cue. They then wrote in "African American."

---

[16] The extent or prevalence of these reactions to the Hispanic origin question among the general population is not known. Our information is based on comments received during cognitive testing, letters to the Census Bureau voicing these concerns and anecdotal information.

[17] Although the combined questions performed as expected in cognitive testing, full scale testing on the RAETT is needed in order to determine if good reporting can be obtained. This will be determined, in part, by comparing the race and origin distribution obtained by the combined questions with the race distribution obtained by the control panel.

We interviewed a number of respondents from the West Indies and Africa. These respondents had a positive reaction to the combined question since many do not identify with "African American" and welcomed the opportunity to specify their national origins.

Coding Race and Ethnicity from the Combined Question

The combined question has only one write-in line for the category "American Indian or Alaska Native" in Part A. Information on Hispanic origin subgroups and Asian and Pacific Islander subgroups as well as write-in entries for the "Some other race" and "Multiracial or biracial" categories should be provided by the respondent in the two write-in lines contained in Part B of the question. As we noted earlier, respondents were able to do this in our cognitive interviews. However, some African-Americans and American Indians commented that the information elicited by Part B is redundant, but these respondents were, nonetheless, able to provide a write-in response.

A significant consideration with the combined question is the coding of the write-in entries in Part B. Race and ethnic origin coding procedures that take into account the unique features of the combined question need to be developed. Another key consideration with the combined question is the extent to which the write-in lines in Part B provide the desired subgroup information on Hispanics and Asian or Pacific Islanders. These considerations cannot be explored with cognitive research techniques. Full scale testing is needed to address these issues. These and other coding and tabulation issues will be examined in the evaluation of RAETT mail return data.

VII CONCLUSIONS

Developing race and ethnic origin questions for the decennial census entails a series of compromises and trade offs. These include technical constraints on the census form, the relatively short-time frame provided to conduct the needed research and diverse and often competing suggestions offered by advisory committees, stakeholders and the general public. Moreover, lack of space on the census form was also a constraint for our research. Some of our design ideas were not feasible because their implementation required more space on the census form than available. Further, a number of our design ideas were not in accordance with the Census Bureau's plan for the Year 2000 Census. This plan calls for making the census form simple and easy to fill out.[18] Thus our questions and format were required to be user-friendly. This criterion ruled out experimentation with elaborate questions or sets of instructions.

As noted, time constraint was also an issue. The development of race and ethnic origin questions for the NCS and RAETT was more complex and took more time and resources than anticipated. We were working with a time schedule which called for the accelerated development of these questions in order to meet critical dates for the OMB review of Directive No. 15,

---

[18] For details on the Census Bureau plans for the Year 2000 Census see U.S. Census Bureau 1996.

congressional deadlines for content items for the 2000 census, and crucial milestones inherent in mounting large scale and complex tests such as the NCS and RAETT.

Based on our research, we believe that there is no perfect or ideal way to ask survey respondents to report race and ethnic origin. In fact, as we have learned, a question which serves the needs of one segment of the population may not include the best way to word or format the question for another group in the population. To some extent, this arises from conceptual differences. For example, making "Hispanic" a race category works well for many Hispanic groups, but not for Cubans; the combination of race and ancestry makes sense for many ethnic groups, but less so for African Americans; Whites and African Americans tend to think of race in terms of color, while Asians tend to think of it in terms of national origins. Sometimes what is best for one group's reporting may interfere with another group's reporting. For example, the inclusion of the write-in line under the American Indian category for the reporting of the enrolled or principal tribe visually breaks the question into two parts and makes it difficult for some respondents to find answer categories below the write-in box. These and other competing factors make the development of race and ethnic origin questions a challenging and difficult task.

An important insight uncovered by our research is that respondents' prior experience with race and ethnic origin questions in surveys (and in other documents such as school forms) is an important determinant of how they interpret and respond to race and ethnic origin questions in self-administered survey and censuses. Thus, most respondents approach our questions with a strong habit of response. For this reason, unexpected modifications such as, changing the instructions in the race question to allow respondents to check more than one racial origin, will go unnoticed by some respondents. However, if this, or other somewhat unexpected changes in race and ethnic reporting (e.g., the use of a multiracial category), are routinely included on survey and census questionnaires, then these methods of reporting will become institutionalized or accepted over time. This institutionalization will lead to greater awareness and use of the new reporting feature among the general population. Therefore, for example, if race questions eliciting more than one racial or ethnic background are routinely included on survey and census questionnaires, the population identifying with more than one race or origin will more than likely increase over time, demographic changes notwithstanding.

R E F E R E N C E S

Bates, Nancy, E. A. Martin, T. J. DeMaio and M. de la Puente (1996). "Context Effects on Factual Questions: Race and Spanish Origin." Forthcoming in Journal of Official Statistics.

Bates, Nancy, M. de la Puente, T. J. DeMaio and E. A. Martin (1994). "Research on Race and Ethnicity: Results From Questionnaire Design Tests." Proceedings of the Bureau of the Census' 1994 Annual Research Conference. Rosslyn, Virginia,
pp. 107 to 136.

Bennett, Claudette, N. McKenney and R. Harrison (1995). "Racial Classification Issues Concerning Children in Mixed-Race Households." Paper presented at the Annual Meeting of the Population Association of America, San Francisco, CA. April 6-8.

Cresce, Arthur A., S. J. Lapham and S. J. Rolark (1992). "Preliminary Evaluation of Data From the Race and Ethnic Origin Questions in the 1990 Census." Paper presented at the Annual Meeting of the American Statistical Association, Boston, Massachusetts, August 10-13.

Crews, D. E. and J. R. Bindon (1991). "Ethnicity as a Taxonomic Tool in Biomedical Research." Ethnicity and Disease, 1, 42-49.

de la Puente, M. and R. McKay (1995). "Research Improves Questions." Civil Rights Journal, Vol. 1, No. 1.

de la Puente, M. and R. McKay (1995b). "Developing and Testing Race and Ethnic Origin Questions for the Current Population Survey Supplement on Race and Ethnic Origin," Proceedings of the 1994 Annual Meeting of the American Statistical Association: Section on Survey Research Methods, Vol. I.

Edmonston, Barry, J. Goldstein and J. T. Lott, eds. (1996). Spotlight on Heterogeneity: The Federal Standards for Racial and Ethnic Classification, Summary of a Workshop. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press.

Edmonston, Barry and J. S. Passel, eds. (1994). Immigration and Ethnicity: The Integration of America's Newest Arrivals. Washington, D.C.: The Urban Institute Press.

Evinger, Suzann (1995). "How Shall We Measure Our Nation's Diversity?" Chance, Vol. 8, No. 1.

Farley, Reynolds (1993). "Questions About Race, Spanish Origin and Ancestry: Findings From the Census of 1990 and Proposals for the Census of 2000." Testimony presented at the hearings of the Subcommittee on Census, Statistics and Postal Personnel of the Committee on Post Office and Civil Service of the U.S. House of Representatives, April 14.

Farley, Reynolds (1990). "Race and Ethnicity in the U.S. Census: An Evaluation of the 1980 Ancestry Question." Paper prepared for the American Statistical Association and the National Science Foundation.

Forsyth, B. and J. Lessler (1991). "Cognitive Laboratory Methods: A Taxonomy." In Biemer, Groves, Lyberg, Mathiowetz and Sudman, eds., Measurement Error in Surveys, New York, Wiley.

Gould, S.J. (1981). The Mismeasure of Man, New York, Norton.

Research Working Group (1995). Research Agenda for the Review of the Racial and Ethnic Categories in Directive No. 15. Interagency Committee for the Review of the Racial and Ethnic Standards.

Kissam, Edward, E. Herrera and J. M. Nakamoto (1993). Hispanic Response to Census Enumeration: Forms and Procedures. Submitted to the U.S. Census Bureau under Task Order No. 46-YABC-2-66027.

McKay, Ruth and M. de la Puente (1995a). "Cognitive Research in Designing the CPS Supplement on Race and Ethnicity," Proceedings of the Bureau of the Census' 1995 Annual Research Conference. Rosslyn, Virginia, pp. 435-445.

McKenney, Nampeo R., C. Bennett, M. de la Puente and R. Harrison (1996). "Census 2000 Research on Race and Ethnicity." Paper presented at the 1996 Joint Statistical Meeting of the American Statistical Association, Chicago, Illinois, August.

McKenney, Nampeo R., C. Bennett, R. Harrison and J. del Pinal (1993). "Evaluating Racial and Ethnic Reporting in the 1990 Census." Paper presented at the 1993 Joint Statistical Meeting of the American Statistical Association, San Francisco, California, August.

McKenney, Nampeo R. and A. Cresce (1992). "Measurement of Ethnicity in the United States: Experiences of the U.S. Census Bureau." Challenges of Measuring an Ethnic World: Science, Politics and Reality. Proceedings of the Joint Canada-United States Conference on the Measurement of Ethnicity, April 1-3,
pp. 173-223.

National Research Council (1994). Proceedings From Workshop on Race and Ethnicity Classification: As Assessment of the Federal Standard for Race and Ethnicity Classification. Committee on National Statistics, National Academy of Sciences, February 1994.

Office of Management and Budget (1995). Federal Register, Vol. 60, No. 166, August 28, 1995.

Office of Management and Budget (1994). Federal Register, Vol. 59, No. 123 1994.

Rodriguez, Clara, E. (1994). "Challenges and Emerging Issues: Race and Ethnic Identity Among Latinos." Proceedings From Workshop on Race and Ethnicity Classification: As Assessment of the Federal Standard for Race and Ethnicity Classification. Committee on National Statistics, National Academy of Sciences, February 1994.

Rodriguez, Clara, E. (1992). "Race, Culture and Latino 'Otherness' in the 1980 Census." Social Science Quarterly, 1992, 73 (4), 930-937.

Tucker, Clyde, R. McKay, B. Kojetin, R. Harrison, M. de la Puente, L. Stinson and E. Robison (1996). Testing Methods of Collecting Racial and Ethnic Information: Results of the Current Population Survey Supplement on Race and Ethnicity. Bureau of Labor Statistics Statistical Notes, No. 40.

U.S. Department of Labor (1995). "A CPS Supplement for Testing Methods of Collecting Racial and Ethnic Information: May 1995." Press release USDL 95-428, Bureau of Labor Statistics.

U.S. Census Bureau (1996). The Plan for Census 2000. April 5, 1996.

U.S. Census Bureau (1992a). Census Questionnaire Content, 1990. CQC-7. Hispanic Origin.

U.S. Census Bureau (1992b). Census Questionnaire Content, 1990. CQC-4. Race.

U.S. General Accounting Office (1993). Census Reform: Early Outreach and Decisions Needed on Race and Ethnic Questions. GAO/GGD-93-36.